**Weather and COVID-19 Deaths during the Stay-at-Home Order in the U.S.**

**eSupplement**

eMethods

 Data Sources

 Statistical Modeling

eReferences

eTable 1

eFigure 1

eFigure 2

eFigure 3

eFigure 4

eFigure 5

eFigure 6

eFigure 7

**eMethods**

**Data Sources**

The data files were joined at the county level and included information on county-specific COVID-19 cases, deaths, demographics, social distancing policies measured at the state-level, weather, ozone, PM2.5, and U.V. information for each county-day from January 1 to June 30, 2020 (eFigure 1). Daily county-level COVID-19 deaths data were collected from state and local public health departments, compiled by the New York Times.1 The New York Times daily county-level COVID-19 deaths data were extracted from the beginning of the pandemic in the U.S. until June 30, 2020, for 3,058 counties with at least one reported COVID-19 case. The COVID-19 dataset was supplemented with county population characteristics from the U.S. Census Bureau.2 The Bureau’s latest (2018) estimations of the U.S. counties’ population characteristics were used. Data compilation included four state-level social distancing measures: when a stay-at-home order was issued, gatherings of 500 or more were banned, public schools were closed, restaurants, gyms, entertainment facilities were closed (eFigure 1).3

County-level COVID-19 data were geocoded using the U.S. county centroid latitude and longitude provided by the National Oceanic and Atmospheric Administration (NOAA)’s National Weather Service.4 The geocoded dataset was linked to NOAA’s Global Historical Climatology Network (GHCN)-Daily temperature data (daily maximum and minimum) and precipitation from 24,854 US weather stations from January 1 to June 30, 2020 (eFigure 1).5 Due to many missing values, additional meteorological information (e.g., average daily temperature, potential for evaporation, wind speed, wind direction, and cloudiness) was not used. Each county-day was assigned with minimum and maximum daily temperatures and precipitation information recorded by the first to the tenth nearest station to the county’s centroid. More than one station was assigned to a county-day to replace missing values of the first nearest station with the values recorded by the next nearest station. A maximum station-to-centroid distance was considered, and the results were tested against changing the maximum distance.

Air pollution information was obtained from the U.S. EPA’s.6 Ozone (O3) and PM2.5 (fine particulate matter) from 2020 were selected because they had substantially fewer missing values than other pollutants and published literature has reported on them in the context of SARS-CoV-2.7–10 Daily maximum 8-hour ozone concentrations (in parts per million, ppm) were reported by 1,218 stations across the U.S. from January 1 to June 30, 2020. Daily average PM2.5 concentrations (in micrograms per cubic meter, μg/m3) were reported by 1,136 stations across the U.S. during the same period (eFigure 1). Each county-day was assigned ozone and PM2.5 information recorded by the first to the tenth nearest air quality monitors to the county’s centroid. Multiple monitor recordings were assigned to a county in order to estimate the nearest station’s missing values with the values recorded by the nearby stations. Lastly, the U.V. light index for the U.S. county centroids came from [www.openweather.com](http://www.openweather.com), using the Python “pyowm” package.11

**Statistical Modeling**

The distribution of new deaths was left-skewed because there were no new COVID-19 deaths in 76% of the county-days and less than two new deaths in 87% of county-days (eTable 1, eFigure 3)―because approximately 85% of the U.S. land area is rural, and rural areas were minimally affected during the spring as the virus was concentrated in urban or suburban areas. The logarithm transformation of per capita new death counts was used as the dependent variable in this analysis. Per capita was defined as the daily death rate by the size of the population over 18 years old in the county. The population below age 18 was excluded due to less frequent COVID-19 deaths in younger age cohorts. A constant (+1.0) was added to daily death rate before dividing by population and taking the logarithm of the quotient to avoid an undefined term (i.e., the logarithm of zero).

Analysis evaluated four statistical models to assess the relationship between COVID-19 deaths and minimum and maximum daily temperatures. The first model was:

|  |  |
| --- | --- |
| $$y\_{ct}=β\_{min}∙minT\_{\left(t-18,t-22\right),c}+β\_{max}∙maxT\_{\left(t-18,t-22\right),c}+δ\_{c}+γ\_{t}+\sum\_{j}^{}π\_{j}X\_{j,st}+ε\_{ct}$$ | (1) |

where the variable $y\_{ct}$ is the log of daily death rate per capita in county $c$ on day $t$. Since county-level daily death rates were weighted by county adult population, the estimated association can also be interpreted as the increase in the number of COVID-19 deaths per adult population. The published literature shows that, on average, infection occurs about 20 days before death.12–19 A five-day window around 20 days before death (i.e., 18 to 22 days before death) was the range used to capture the time window when infection began. The variables $minT\_{(t-18,t-22),c}$ and $maxT\_{(t-18,t-22),c}$ are the average minimum and maximum temperatures in the five-day window.

The set of county fixed-effects is given by $δ\_{c}$. These are dummy variables that represent each county in the data. These fixed-effects control for factors that are constant within counties over the period of this study but can vary across counties. They include factors such as populations density, age, gender, education, ethnic distribution, and other demographic characteristics of a county, median income, income distribution, industry composition, labor factors and all other economic characteristics of a county, health care features and resources of a county, cultural factors in a county, such as social norms, attitudes, and aspirations, geography and climate of a county, political characteristics, and infrastructure of a county, to name a few.

The set of time fixed-effects for each day is given by $γ\_{t}$. 103 dummy variables were included in the regression to represent every 104 days in the analysis sample. The $γ\_{t}$ fixed-effects pick up factors that are constant across counties but can vary daily. Federal government policies or anything that affects the whole nation, such as the Center for Disease Control’s 6-foot physical distance measures, travel restrictions, and immigration policies, and national or global news events, are examples of effects fixed across time at the U.S. level.

To capture differing policies and procedures in response to COVID-19 that occurred at the state level,$ X\_{j,st}$ is a set of four policy-response dummies that control for four state-level policy changes over time: (1) stay-at-home order, (2) order of no gathering of more than 500 people, (3) order of public school closures, and (4) order of closure of restaurants, entertainment venues, and gyms. In other words, for each policy, a dummy variable was equal to one for a county on a specific day if the state to which the county belongs had that policy in place on that day. The averages of these dummy variables 18 through 22 days before death were included as independent variables.

To strengthen modeling, Model (2) was defined to account for county-level time-trends, $θ\_{c}t$:

|  |  |
| --- | --- |
| $$y\_{ct}=β\_{min}∙minT\_{\left(t-18,t-22\right),c}+β\_{max}∙maxT\_{\left(t-18,t-22\right),c}+δ\_{c}+γ\_{t}+θ\_{c}t+\sum\_{j}^{}π\_{j}X\_{j,st}+ε\_{ct}$$ | (2) |

These time-trends control for tendencies in the log of per capita daily death rate specific to each county ($y\_{ct}$). In total, 1,322 additional independent variables were included in the regression.

Models (1) and (2) include daily time fixed-effects, $γ\_{t}$, controlling for national-level factors (e.g., federal-level policies or national news events) that may change daily but are constant across all counties on a specific day. Regional-level day fixed-effects were added to control for additional time-varying factors unique to a certain U.S. region. That is, regional-level day fixed-effects control for factors that (1) can vary day by day, (2) are constant for counties in a certain region on a specific day, but (3) vary in other region’s counties on the same day. For example, some neighboring states in certain regions coordinated COVID-19 responses. Such coordination means that each day these states implement policies that are similar in that region but differ from other regions. Region-specific day fixed-effects are represented by $μ\_{rt}$ in Models (3) and (4). regions used in this study were New England, Mid-East, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain, and Far West.20

|  |  |
| --- | --- |
| $$y\_{ct}=β\_{min}∙minT\_{\left(t-18,t-22\right),c}+β\_{max}∙maxT\_{\left(t-18,t-22\right),c}+δ\_{c}+μ\_{rt}+\sum\_{j}^{}π\_{j}X\_{j,st}+ε\_{ct}$$ | (3) |

The number of dummy variables representing $μ\_{rt}$ fixed-effects in these models were calculated as 8 regions for 104 days totaled 832 variables. In Model (4), county-level time-trends were added to Model (3):

|  |  |
| --- | --- |
| $$y\_{ct}=β\_{min}∙minT\_{\left(t-18,t-22\right),c}+β\_{max}∙maxT\_{\left(t-18,t-22\right),c}+δ\_{c}+μ\_{rt}+θ\_{c}t+\sum\_{j}^{}π\_{j}X\_{j,st}+ε\_{ct}$$ | (4) |

Model (4) provides the strongest specification among the four models and controls for numerous time-constant and time-varying factors. It includes 3,484 variables other than the minimum and maximum daily temperatures, some of which control for multiple factors by themselves (e.g., county fixed-effects and day-fixed effects). In this model, any remained confounder should vary both by space (counties) and over time (days). Confounder(s) should be correlated with both daily death rate (dependent variable) and temperature variables (the independent variables of interest). Time-varying confounders that may be correlated with both daily death rate and temperature variables (e.g., precipitation, pollution, and U.V. light) were added as independent variables to Model (4). Estimates of the association between precipitation, pollution and U.V. index with COVID-19 fatality rates were obtained by adding precipitation, pollution, and U.V. light daily average between days 18 through 22 before death in regressions. Model (4) was estimated for various ranges of temperature to correct for potential nonlinearity of the association of COVID-19 deaths and minimum and maximum daily temperatures.

A robust solution to the problem of serial correlation in panel data estimation that corrects the standard errors of the coefficients is to cluster them at the cross-sectional unit in the data (i.e., counties).21–24 As a result, lagged dependent variable was not needed as a regressor. This method of clustering standard errors at the county level accounts for the serially correlated structure of the standard errors within each county.

The described statistical modeling approach has several strengths. It more deliberately controlled for confounding factors possible through the use of county fixed-effects and time fixed-effects. One benefit of county fixed-effects is that they adjust estimates for any confounder that is constant over the period of this study but can vary across counties, even if data on the factor are not available or the factor is not measurable. County-level fixed-effects include factors such as population density, distribution of age, gender, education, and ethnicity, as well as other demographic characteristics of a county, income distribution, industry composition, job distribution, and all other time-constant economic characteristics of a county (e.g., health care quality, county resources, cultural factors, and political characteristics). Time fixed-effects captured factors affecting the variation in daily death rate but were constant across counties and regions in the United States, such as national or global responses (e.g., travel restrictions and immigration restrictions). They do not require a functional form to be specified to capture changes in policies. County time trends control for any trend(s) in death rate within each county separately.

**eReferences**

1. Smith M, Yourish K, Almukhtar S, Collins K, Ivory D, Harmon A. Coronavirus (Covid-19) Data in the United States [database online]. The New York Times. Published 2020. Accessed May 18, 2020. https://github.com/nytimes/covid-19-data

2. U.S. Census Bureau. County population by characteristics: 2018 [database online]. Published 2020. Accessed May 20, 2020. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html.54.

3. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong social distancing measures in the United States reduced the COVID-19 growth rate. Health Aff. 2020:10.1377/hlthaff. doi:10.1377/hlthaff.2020.00608

4. National Weather Service. Counties of the U.S. used by National Weather Service to issue county based forecasts and warnings [database online]. Published 2020. Accessed May 19, 2020. https://www.weather.gov/gis/Counties.

5. Global Historical Climatology Network. Daily temperature and precipitation reports data tables [database online]. Published 2020. Accessed May 20, 2020. https://www.climate.gov/maps-data/dataset/daily-temperature-and-precipitation-reports-data-tables.

6. United States Environmental Protection Agency. Outdoor air quality data [database online]. Published 2020. Accessed June 4, 2020. https://www.epa.gov/outdoor-air-quality-data/download-daily-data.

7. Setti L, Passarini F, De Gennaro G, et al. Searching for SARS-COV-2 on particulate matter: A possible early indicator of COVID-19 epidemic recurrence. Int J Environ Res Public Health. 2020;17(9). doi:10.3390/ijerph17092986

8. Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. medRxiv. 2020. doi:10.1101/2020.04.05.20054502

9. Blanchard EL, Lawrence JD, Noble JA, et al. Enveloped virus inactivation on personal protective equipment by exposure to ozone. medRxiv. 2020. doi:10.1101/2020.05.23.20111435

10. Dubuis M-E, Dumont-Leblond N, Laliberté C, et al. Ozone efficacy for the control of airborne viruses: Bacteriophage and norovirus models. PLoS One. 2020;15(4):e0231164. doi:10.1371/journal.pone.0231164

11. MIT Open Weather Map. Python-Pyowm Package [software program]. Accessed June 8, 2020. https://pypi.org/project/pyowm/.

12. Jung S-M, Akhmetzhanov AR, Hayashi K, et al. Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference using exported cases. J Clin Med. 2020;9(2). doi:10.3390/jcm9020523

13. Linton NM, Kobayashi T, Yang Y, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. J Clin Med. 2020;9(2):538. doi:10.3390/jcm9020538

14. Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect Dis. 2020;20(7):793-802. doi:10.1016/S1473-3099(20)30230-9

15. Tindale L, Coombe M, Stockdale JE, et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. medRxiv. 2020;2020.03.03.20029983. doi:10.1101/2020.03.03.20029983

16. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis J. 2020;26(7). doi:10.3201/eid2607.200282

17. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. The novel coronavirus, 2019-nCoV, is highly contagious and more infectious than initially estimated. medRxiv. 2020;2020.02.07.20021154. doi:10.1101/2020.02.07.20021154

18. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. Ann Intern Med. 2020;172(9):577-582. doi:10.7326/M20-0504

19. Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis. 2020;20(5):553-558. doi:10.1016/S1473-3099(20)30144-4

20. Bureau of Economic Analysis. BEA Regions [database online]. 2020. Accessed June 8, 2020. https://apps.bea.gov/regional/docs/regions.cfm

21. Arellano M. Panel Data Econometrics. Oxford, UK: Oxford University Press; 2003.

22. Hansen CB. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. J Econom. 2007;141(2):597-620. doi:https://doi.org/10.1016/j.jeconom.2006.10.009

23. Stock JH, Watson MW. Heteroskedasticity-robust standard errors for fixed effects panel data regression. Econometrica. 2008;76(1):155-174. doi:10.1111/j.0012-9682.2008.00821.x

24. Cameron AC, Miller DL. A practitioner’s guide to cluster-robust inference. J Hum Resour. 2015;50(2):317-372. doi:10.3368/jhr.50.2.317

**eTable 1. Summary statistics of COVID-19 death and weather**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Standard** |  | **25th** |  | **75th** | **95th** |  |
| **Variables** | **Mean** | **Deviation** | **Minimum** | **Quantile** | **Median** | **Quantile** | **Quantile** | **Maximum** |
| Number of New Deaths | 1.1 | 5.2 | 0.0 | 0.0 | 0.0 | 0.0 | 5.0 | 258.0 |
| log(Daily Deaths per 18+ Population) | 2.2 | 4.0 | 0.0 | 0.0 | 0.0 | 0.0 | 10.3 | 13.8 |
| Average Minimum Temperature (F) | 43.9 | 10.9 | -0.1 | 35.6 | 43.4 | 51.8 | 62.8 | 77.8 |
| Average Maximum Temperature (F) | 65.4 | 11.9 | 26.8 | 56.2 | 65.8 | 74.4 | 84.0 | 108.2 |
| Average Precipitation (mm) | 2.9 | 1.5 | 0.0 | 1.9 | 3.2 | 4.0 | 4.9 | 6.3 |
| Average Ozone Concentration (ppb) | 41.2 | 5.7 | 1.2 | 38.2 | 41.8 | 44.8 | 49.2 | 90.8 |
| Average PM2.5 Concentration (μg/m3) | 6.6 | 2.8 | 0.1 | 4.7 | 6.2 | 8.0 | 11.6 | 37.1 |
| Average UV Light Index | 7.1 | 1.8 | 2.1 | 5.8 | 7.1 | 8.5 | 9.9 | 12.0 |

These statistics include air quality indicators 18 to 22 days before death to COVID-19 in 1,323 counties and 59,990 county-days of the analysis sample

**eFigure 1. Data assembling procedure**

County-Day Data File 1:

* COVID-19 Cases and Deaths ([NYtimes](https://github.com/nytimes/covid-19-data))

County-Day Data File 2:

* COVID-19 Cases and Deaths
* County Centroid Geo Info

County Centroid Geo Info ([National Weather Service, NOAA](https://www.weather.gov/gis/Counties))

County-Day Data File 3:

* COVID-19 Cases and Deaths
* County Centroid Geo Info
* Demographics

County population Characteristics, 2018 Estimation ([Census Bureau](https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html))

County-Day Data File 4:

* COVID-19 Cases and Deaths
* County Centroid Geo Info
* Demographics
* Social Distancing

State-Level Social Distancing Measures ([Health Affairs](https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2020.00608))

County-Day Data File 5:

* COVID-19 Cases and Deaths
* County Centroid Geo Info
* Demographics
* Social Distancing
* Temperature and Precipitation

Geocoded US Weather Data, 24,854 Stations ([GHCN, NOAA](https://www.climate.gov/maps-data/dataset/daily-temperature-and-precipitation-reports-data-tables))

County-Day Data File 6:

* COVID-19 Cases and Deaths
* County Centroid Geo Info
* Demographics
* Social Distancing
* Temperature and Precipitation
* Ozone and PM2.5

Geocoded Ozone and PM2.5 Data, 1,218 and 1,136 Stations ([EPA](https://www.epa.gov/outdoor-air-quality-data/download-daily-data))

County-Day Data File 7:

* COVID-19 Cases and Deaths
* County Centroid Geo Info
* Demographics
* Social Distancing
* Temperature and Precipitation
* Ozone and PM2.5
* UV Light Index

Geocoded UV Light Data ([OpenWeather.com](https://openweathermap.org/api/uvi))

**eFigure 2. Data refinement procedure**

County-Day Data File 7:

Number of Counties: 3,141

Number of Counties: 571,662

Dates: 1 Jan 2020 to June 30 2020

County-Days with the Next Nearest Weather Station’s Distance from County Centroid − Preceding Weather Station’s Distance from County Centroid > 25 Miles

County-Day Data File 8:

Number of Counties: 3,088

Number of Counties: 562,016

Dates: 1 Jan 2020 to June 30 2020

County-Days with Nearest Ozone-Recording station > 60 Miles away

County-Day Data File 9:

Number of Counties: 2,608

Number of Counties: 443,624

Dates: 1 Jan 2020 to June 30 2020

County-Days with Nearest PM2.5-Recording Station > 60 Miles away

County-Day Data File 10:

Number of Counties: 2,483

Number of Counties: 408,138

Dates: 1 Jan 2020 to June 30 2020

County-Days before the First COVID-19 Death

County-Day Data File 11:

Number of Counties: 1,698

Number of Counties: 124,070

Dates: 29 Feb 2020 to June 30 2020

Limiting the Sample to Shelter-in-Place to Reopening Period

County-Day Data File 12 (Final):

Number of Counties: 1,323

Number of Counties: 59,990

Dates: 29 Feb 2020 to June 30 2020

|  |
| --- |
| **eFigure 3. Histogram of daily death rates** |
|  |  |
| 1. $\frac{D\_{ct}}{18+ pop\_{c}}$
 | 1. $log⁡(\frac{D\_{ct}}{18+ pop\_{c}})$
 |
| The variable $D\_{ct}$ stands for the number of new deaths in county $c$ on day $t$. $18+ pop\_{c}$ is the size of the population over age 18. Figure (a) shows the daily death rate per 18+ population in county $c$ on day $t$. Figure (b) is the logarithm of this measure, used as the dependent variable in this study. |

|  |
| --- |
| **eFigure 4. Histograms of weather and air quality variables** |
|  |  |
| 1. Average Minimum Temperature $minT\_{\left(t-18,t-22\right),c}$
 | 1. Average Maximum Temperature $maxT\_{\left(t-18,t-22\right),c}$
 |
|  |  |
| 1. Average Precipitation
 | 1. Average O3
 |
|  |  |
| 1. Average PM2.5
 | 1. Average UV Index
 |
| All these averages are calculated over a five-day window 18 through 22 days before death. |

**eFigure 5. Percentage change in deaths as predicted by minimum daily temperature**



Presumed Exposure Period: 8 to 12 Days Before Death



Presumed Exposure Period: 28 to 32 Days Before Death

These are percentage (in decimals) changes in daily deaths per 18+ county population for a 1°F increase in the five-day average of minimum daily temperature in days 8 to 12 and days 28 to 32 before death, stratified by county-days based on their highest average minimum temperature.

**eFigure 6. Percentage change in deaths as predicted by maximum daily ozone**



Presumed Exposure Period: 8 to 12 Days Before Death



Presumed Exposure Period: 28 to 32 Days Before Death

These are percentage (in decimals) changes in daily deaths per 18+ county population for a 1 ppb increase in the five-day average of maximum daily ozone level in days 8 to 12 and days 28 to 32 before death, stratified by county-days based on their lowest average maximum daily ozone level.

**eFigure 7. Percentage change in deaths as predicted by maximum daily ozone**



Max. Distance from O3 Monitor: 40 miles

Number of Observations: 11,278−47,845

Number of Counties: 927−1,056



Max. Distance from O3 Monitor: 20 miles

Number of Observations: 5,020−22,606

Number of Counties: 419−474

These are percentage (in decimals) changes in daily deaths per 18+ county population for a 1 ppb increase in the five-day average of maximum daily ozone level in days 18 to 22 before death, stratified by county-days based on their lowest average maximum daily ozone level and by the maximum distance of ozone-recording monitor to the county centroid.